Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2011 Proceedings - All Submissions

8-5-2011

Detecting Community Influence Echelons in Twitter Network

Weiyi Sun
SUNY at Albany, Ws344453@albany.edu

Detecting Community Influence Echelons in Twitter Network

Weiyi Sun
SUNY at Albany
Ws344453@albany.edu

ABSTRACT

We study the interactions in a coherent community on Twitter to examine its structure. In particular we examine if there exists a hierarchical influence structure induced by the interactions which reflect a ranked partition of the users in the community where users retweet (forward) only messages from other users belonging to an equal or higher ranked group. We extract such ranked partition of the community and show it to roughly align with independently constructed influence score of users in each echelon. Our research suggests that the relationship and forwarding behavior in online microblogging community is affected by the underlying social influence structure and the understanding of the structure may help us better predict the information diffusion on such online communities.

Keywords

microblogging, online community, social network analysis, network structure

INTRODUCTION

With computing and networking technology thriving in the past decade, fully-fledged digital devices and ubiquitous network connection give raise to the rapid growth in microblogging and social networking services, such as Twitter.com, Google Buzz and so on. Such online services enable users to easily broadcast short messages anywhere using almost any Internet-enabled devices. The growth and popularity of these services has dramatically changed the way information flows in our life. As a result, journalism, political propagation, and product marketing have thus begun to undergo a tremendous transition. Many a time we find traditional mass media citing Twitter as their information source; we rely on Twitter and other microblogging services to get information about political unrest from local residents when the government cut off formal sources of information; nowadays, we can hardly find any major business that is not presenting itself on Twitter.com. All of these transitions fuel our curiosity about the questions of “how does information diffuse on micorblogging services” and “how do people influence each other on these platforms”. Understanding these questions will help us better design social media, social media campaigns, as well as better comprehend the theories of influence in general.

The theories of influence have evolved over the past several decades. Earlier theories stress on the disproportional influence power of a small number of individuals in a society who are good at convincing people.(Keller et al. 2003) (Rogers 1962) More modern theories argue that the emphasis placed on “influentials” are overrated, and that the network structure among ordinary users and the readiness of a society to adopt a new idea are the key factors that determine influence. (Domingos et al. 2001, Richardson et al. 2002, D. Watts et al. 2007) These theories remain as theories due to lack of empirical data to validate them. The recent growth of microblogging services provides a suitable environment for such empirical studies. Services, such as Twitter.com, provide a vast quantity of recorded information of user’s information sharing behavior, the majority of which is publicly accessible. User’s information behavior leave digital traces on these online platforms that enable researchers to examine real social network at “unprecedented levels of scale and resolution”.(Kleinberg 2008)

In this paper, we observe a semi-hierarchical influence structure in a cohesive Twitter community in terms of the retweet relationship social graph -- the retweet graph of this community is found to be highly acyclic. We extract the partial order in the retweet graph, and show that the resulting echelons matches quite well with a social influence indicator constructed by measures of user influence, including number of friends, hub/authority score of friendship, number of mentions, and number of retweets.

The next session starts with introducing the basic concepts, definitions and background of Twitter.com, and then put forward the network model that we use in this research. After the Data Collection section, we present the indications of the existence of the above described echelon structure and compare it with a social influence score constructed by a number of influence measures. The fact that the derived echelon, to some degree, reflects the users' influence measure, suggests that the retweet behavior is largely governed by the underlying semi-hierarchical social influence structure.

Related Work

Twitter.com has been brought to academic attention shortly after it just began to prosper. Early researches focus on providing descriptive information about this new form of social network. A Java et al(2007) studies the topological and geographical properties of the Twitter social network by collecting and analyzing all the messages posted on Twitter during a month; Huberman et al(2008) examines a dataset of 309,740 users and finds that users pay attention only to a limited subgroup of their declared "friends"; Krishnamurthy el al (2008) collects tweets and user information by sampling the public timeline and active users, and classifies users and their behaviors.

Recently, more researchers start to look at user influence on Twitter.com. (Kwak et al. 2010, Weng et al. 2010, Cha et al. 2010, Bakshy et al. 2011). The results and recommendations are somewhat contradictory. Kwak et al. (2010) compare three different measures of influence (number of followers, PageRank, and number of retweets), finding that PageRank and follower rankings are similar but retweet ranking is quite different. Cha et al. (2010) also compare three measures: number of followers, number of retweets and number of mentions, reporting that the three measures have moderate correlations. Weng et al.(2010) develop a topic sensitive PageRank-like measure, TwitterRank. Bakshy et al. (2011) track the passing of URLs. and use regression tree model to find out that the past performances are the best predictors and the number of followers is also an informative predictor.

Most of the above mentioned researches try to build inclusive datasets – not putting any boundary on their datasets, with the exception of Weng et al. (2010) who collect their dataset within a geological boundary– Singapore. This work differentiates from previous research in that its target dataset is a carefully chosen cohesive sub-community of the Twitter network – a tech enthusiasts' community. Given the drastic heterogeneity in user behavior among Twitter users, the actual dynamics of a certain user subgroup can be completely different from the overall characteristics of entire complex network presented in previous researches. It is common for a complex network to show a locally-focused community structure. (Lancichinetti et al. 2009) In a large social network such as Twitter's, a user is most likely to interact and influence only his/her local community. Observation at the level of the entire network may overlook what is truly going on at the community level. Moreover, instead of selecting arbitrary measure of influence, we focus on the underlying influence structure and use influence measures only as validation and comparison. We are not aware of any other literature that studies the structural property of a cohesive community on Twitter.com.

TWITTER NETWORK MODEL

We define the network model that describes the user relations and behavior ties in the Twitter sub-community in this section, after a background introduction and term definition of Twitter.com

Twitter.com

Twitter is one of the most popular microblogging services. It allows users to post messages, up to 140 characters per massage, known as tweets. Tweets are displayed on the author's profile . By default, all tweets and user profiles are publicly accessible, unless otherwise specified by the user.

Users can subscribe to other twitter users. This subscribing action is called "following". The users that one follows are called one's "friends". Users will automatically receive the tweets posted by their friends. The "friendship" relation needs not to be reciprocal. One can follow any other user as long as such user has a public profile. Users will be notified when they are followed by others, and they may choose whether to follow back or not.

One can mention other users in a tweet, by referring to their screenname with the prefix "@". For example, "meeting iMouse team with @duoshute". Tweets that contain such @-strings are called "mentions". The user being mentioned, or being "replied to", will receive a notification. Mentions allow users to conduct public conversations. By adding the letters "RT" (case-insensitive) in front of an screenname, it is understood as a retweet behavior (forwarded message), e.g. "RT @TheNextWeb: Breaking: Wikileaks is down on verge of massive documents dump". Another symbol that is widely used in tweets is the "#" sign. Beginning a string with "#" sign indicates a tag or topic. For example, "President Bill Clinton speaking at #time100 http://yfrog.com/2mbkzwj". Topic signs make tweets easier to follow.

Social Graph Definitions

In this paper, we use the following graph model to describe the sub-community that we observe and the relational and behavioral ties among members in such sub-community.

On the set of *n* Twitter users in the selected sub-community $U:\{u_i \mid i=1,2,\cdots,n\}$, we define 3 relations:

- R_1 *Friendship relation*: a friendship relation exists from actor u_j to actor u_k if u_j follows u_k on Twitter.
- R_2 *Mention relation*: a mention relation exists from actor u_j to actor u_k if within u_j's latest t_1 tweets. t_1 is a threshold value.
- R_3 *Retweet relation*: a retweet relation exists from actor u_j to actor u_k if within u_j's latest t_2 tweets.

Let directed graph $G_i\left(U,\vec{A}_i\right)$ represent the R_i networks (i=1, 2, 3). Each Twitter user observed in our sub-community is represented by a vertex in G_i. $\vec{A}_i$ is the arc set of G_i, is determined by the R_i relation: an arc exists from vertex u_j to vertex u_k if and only if there is an R_i relation from u_j to u_k. Figure 1 is an illustration of the network model. It depicts some *Friendship*, *Mention* and *Retweet* sub-graphs of our dataset.

Figure 1. A network visualization (ego: the blue node): the yellow edges represent *Friendship* relation; green edges for *Mention* relation and the red edges for *Retweet* relation.

DATA COLLECTION

To examine the relational and behavioral pattern in a coherent sub-community on Twitter.com, we select a specific sub-community and use the Twitter APIs[1] to collect the social relations and interactions among them. For privacy concern, only those who choose to open their profile to the public will be included in the dataset. All the information in this dataset can be obtained from the Twitter website without password protection.

[1] http://apiwiki.twitter.com/Twitter-API-Documentation

The sub-community of choice is a group of popular technical enthusiasts, including technology bloggers, entrepreneurs, ventures capitalists and developers. The reason for selecting this specific sub-community is threefold. Firstly, the members of this community are among the first adopters of the Twitter service. They are most familiar with the features and usages of this service, and usually indicate the future tendency of the Twitter network. Secondly, this is a well-connected and active sub-community, which ensures that the community be active and functional. Lastly, this is a group in which news and information is highly valued and efficiently diffused. The structure of this sub-community may provide us new knowledge of the information flow on Twitter.

Data Collection Method[2]

As the starting point, we choose the accounts listed in the FavStar[3] "Top 50 Tech"[4] list . The "Top 50 Tech" list contains the 48 Twitter accounts that generate the most tweets that are "favorite" by many people. For our data collection purpose, this list approximates the core set of user in the targeted area of technology enthusiast community.

From this set of users, we adopt a snowball data collection strategy – to repeatedly include existing target users' Twitter contacts into our target dataset. The challenge in the snowball data collection strategy is to ensure that the users whom we add to our dataset are actually the ones that are actively involved in this sub-community. Simply including all the declared "friends" of a user (the accounts that the user follows) is problematic because some users follow back everybody who follows him/her, which leads to as many as 300,000 declared friends. Clearly, Not all of these followees fit our data collection purpose in that many of these users don't actually interact with any of the users in our dataset and thus do not belong in the sub-community. To keep our target user group cohesive, we collect the maximum of 400 friends from each user, and put together the base user set of 2,568 accounts[5]. We impose two additional criteria to further ensure the activeness and coherence of the final selected cohort, that is, the user should either get "voted" as friend by multiple different base user set members (to reduce the chance of accidentally including an outsider) or have been involved in conversations with members from the selected group. After testing different thresholds, we decide to select users who are friends of at least 45 (2%) other accounts in the base user set (as this cut-off value provides a reasonable sized user set for the purpose of this study) and those who has been mentioned by at least 2 members of the selected group. The final use set consists of 775 users. We retrieved the latest 200 tweets of these users140,678 tweets in total, 72,870 of which are *mention*s, among which 10,609 tweets are *retweet*s.

Dataset Overview

The final dataset consists of a user-attribute table of the 775 target users and the 3 directional relations among them: *Friendship*, *Mention* and *Retweet*.

The user-attribute table contains profile information of the users, including their Twitter User ID, Name, Twitter ScreenName, Location, Description, URL, total number of followers on Twitter, total number of friends, the account's date of creation, the number of tweets that she favorites, and the total number of tweets that she ever posted. Geographically, the largest subgroups of the dataset are from California (around 25%, 65% of which are from San Francisco), following by New York (about 10%). The average number of followers of the dataset is 163,916.

The network statistics in Table 1 below is to provide some statistics of the dataset. Noticing that the *Retweet* graph have a large portion of isolate nodes, we also provide the density of the linked components of the *Retweet* graph.

Density (Faust. 1994) is defined as the number of edges of each graph divided by the number of possible edges in the graph:

$$D_i = {|A_i|} \big/ {|U|(|U|-1)}$$

As described earlier, the *Friendship* relation is the most common relation in this data set, and has a density of 0.0307. It is much higher compared to the reported 0.000107 density in previous literature (Java et al. 2007), because of the user selection criteria that we use here. The *Mention* relation occurs about half as often as the *Friendship* relation. A user can mention another user whether the other user is his friend or not. However, most of the *Mention*s happens between friends(Huberman et al. 2008). The *Retweet* relation observed in this dataset is rather sparse.

---

[2] The Data collection of this research was completed on May 5, 2010. Final data set is available at http://sunweiyi.com/TwitterInfluence/
[3] http:// favstar.fm Favstar is a service that tracks the most favorite tweets on Twitter
[4] The current list can be found at: http://listorious.com/Favstar/top-50-tech The list that this research used can be found in the data packet.
[5] It should be noted that the number suggests that there is significant overlap in the friendship relation in this selected group.

Reciprocity is the ratio of the number of reciprocal edges over the total edge count. The reciprocity of *Friendship* and *Mention* relation in this dataset are around 0.21, which is very close to the 0.22 *Friendship* reciprocity reported in (Kwak et al. 2010) The *Retweet* relation displays the least reciprocity. The one-sidedness of these relations motivates us to look at directedness of the ties beyond dyad level -- at a network level.

	Friend (G1)	Mention (G2)	Retweet (G3)
Density	0.0307	0.0154	0.0027 [linked: 0.0046]
Reciprocity	0.2144	0.2128	0.0455
|Ei|	18428	9261	1607
Number of Isolates	1	9	205

Table 1. Basic network statistics

SOCIAL ECHELONS DETECTION

Indication of Partially Graded Structure

In his book chapter(Krackhardt 1994), Krackhardt develops a method for measuring the degree to which a social network displays a hierarchical structure. He compares four graph theoretic measures of the social network to a pure hierarchy structure – a directed tree. The four measures he chooses are connectivity, graph hierarchy, graph efficiency and least-upper-boundedness. They are necessary and jointly sufficient conditions for a graph to be a directed tree.

Nonetheless, it is hardly possible for an informal organization to resemble a strict hierarchical structural, because it is usually made up by smaller local communities which cause overlapping smaller structures. To allow overlapping structures, we will instead examine whether the network resembles a graded graph. Here, a grading of a directed graph $\vec{G}(V,\vec{A})$, with vertex set V and arc set $\vec{A}$, is defined as a partitioning of V into $V_1,V_2,\cdots,V_k$ such that if $\overrightarrow{xy}\in\vec{A}$, then $x\in V_i$ and $y\in V_j$,where $i<j$ for some i, j. (Bollobas 1998)

Clearly the necessary and sufficient condition for a graph to have a partial grading, is that the directed graph is acyclic. That is, if there is a directed path from x_0 to x_1 in a graph $\vec{G}$, then a directed path from x_1 to x_0 doesn't exist in $\vec{G}$. The requirement is equivalent to what Krackhardt defines as "graph hierarchical" index. Thus, we will use Krackhardt Hierarchy index to measure the degree to which a social network tends to be partially graded.

To compute the Krackhardt Hierarchy score of a directed graph $\vec{G}(V,\vec{A})$, we first get the reachability graph($\vec{G}_R$) of $\vec{G}$: $\vec{G}_R(V,\vec{A}_R)$,where $\overrightarrow{xy}\in\vec{A}_R$ iff a directed path $\vec{P}$ exists in $\vec{G}$, whose head and tail are x and y respectively ($x,y\in V$). The Krackhardt Hierarchy score is defined as the percentage of unsymmetrical ties in the reachability graph:.

$$H=\frac{\left|\left\{\overrightarrow{xy}\,\middle|\,\overrightarrow{xy}\in\vec{A}_R,\overrightarrow{yx}\notin\vec{A}_R\right\}\right|}{\left|\vec{A}_R\right|}$$

Hence, H is a value in [0, 1] that indicates the "acyclic" level of the original directed graph $\vec{G}$. A strongly connected cycle has the H score of zero because each pair of vertices is strongly connected in both direction, whereas a directed tree has the H score of 1 in that it is acyclic and each directed path is not reversible.

For each of the three relations, we calculate the Krackhardt Hierarchy score (Table 2). To understand the average expected Krackhardt Hierarchy value in a network of similar size and order, we also computed the Krackhardt Hierarchy score of 50 775-vertex random graphs of size 1607, 9261 and 18428. The average H score for a network of the same size as Retweet network is around 0.3895(Standard deviation 0.032). For networks of similar size as the Mention and Friendship networks, the average Krackhardt Hierarchy scores are 0.001 and 0.000 respectively with standard deviation<< 0.0000. Thus, we find that all three networks have high Krackhardt hierarchy scores, especially the Friendship and Retweet network.

	Friend (G1)	Mention (G2)	Retweet (G3)
H Score	0.6411	0.1363	0.7912

Table 2. Krackhardt Hierarchy Score

As demonstrated above, high Hierarchy score infers the approximation towards a partial grading. In the Twitter community social network, the occurrence of the partial grading is likely to suggest that the relations are determined by social status, prominence and influence. Prominent users are more likely to be followed by but no follow back less prominent users; users are more likely to retweet those who are more preeminent. On the other hand, Twitter users use the "@" mention action to carry on public conversations. The relatively low Krackhardt Hierarchy score of the *Mention* network (compared to the other two networks) suggests that in this organic community, members seek to communicate with each other following a less hierarchical structure.

Obtaining DAG

As shown above, *Retweet* is a highly directed behavior. Only 4.6% of the *retweet* relation is reciprocal. In another word, if one has retweeted other people in their most recent 200 tweets, on average, only 4.6% of them have ever retweeted one back in any of their latest 200 tweets. Moreover, only 20.88% of the paths in the *Retweet* relation graph that are cyclic, which suggests that this graph is quite "close" to an acyclic graph – the equivalent representation of a partial order. In order to find such a partial order, we need to convert the graph to a DAG.

There are several ways to convert a regular directed graph to DAG. A well-known solution is by solving the minimum feedback arc set problem. Given a weighted directed graph $\vec{G}(V,\vec{A})$, the minimum feedback arc set problem consists of finding a minimum weight set of arcs $\vec{A}' \subseteq \vec{A}$ such that the directed graph $\vec{G}'(V,\vec{A} \setminus \vec{A}')$ is acyclic. (Karp 1972) However, in our case, it is difficult to define what is to be minimized, because each retweet bears different level of significance in determining the "prominence" of the sender and receiver, but the *retweet* graph only contain information about retweet frequency, and there is no information about how "important" each retweet is. Besides, it is not preferable if the final echelons obtained by the DAG are sensitive to the selection of the arc set.

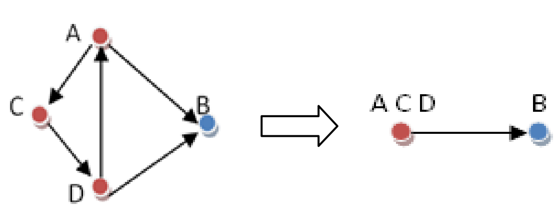

Figure 2. Example relation

To avoid arbitrary removal of links, and to ensure that the DAG outcome is unique and reliable, we use a different strategy to obtain a DAG, by grouping strong components. A strong component of a directed graph G is a maximal strongly connected6 induced subdigraph of G. All the vertices in a strong component are reachable to each other. Figure 2 shows a

[6] A directed graph G is strongly connected if for any two vertices of G, there exists a directed closed walk containing both of them.

digraph in which node A, C and D form a strong component. If this were a subgraph of the Retweet network, it would mean that A retweeted C, and C retweeted D, and D retweeted A, which put all of them on equal grounds. In this sub-community, we would think that A, C, D are in the same echelon, while B is probably "superior" to them because this strong component forwards B's messages, but B has never forwarded any of their messages. If we treat node A, C and D as one entity, (see Figure 3), we can obtain a simplified digraph with a clear order. In short, the basic idea of this strategy is to find strong components in the *Retweet* graph, and treat each strong component as a new entity, thus reducing reciprocal arcs. The underlying assumption is that users who retweet each other belong to the same echelon.

The algorithm is described below. The basic idea is to find strong components in the *Retweet* graph, treat each strong component as a new entity, thus reducing reciprocal arcs, and repeat this process until a DAG is obtained. When building blockmodel, we use the α criteria. (Wasserman et al. 1994) Since we usually cannot expect all the users in a block to be structurally equivalent actors – to have blocks of all 0's or positive values, it is common to compare the block density (δ) with the overall density (α) to determine whether the block will take 0 or positive value:

$$bl = \begin{cases} 0 & \delta < \alpha \\ \sum\limits_{a \in block} a & \delta \geq \alpha \end{cases}$$

```
Algorithm GetDAG(G=(V, A))
   Begin
       While ( G is not acyclic)
       Begin
          SC<-(empty, empty)
          SCSet<-empty
          While  (G contains a strong component)    %Phase 1: get all the strong components
          Begin
              Let SCC be a strong component in G.
              G<-G\SC
              add SC to SCSet                        % set SCSet stores all the SC subgraphs
          End
          V'<-{All the singleton vertices in G, and all the strong components(each component is treated
          as one new vertex)                                                    %  Phase  2:  create
          blockage graph
          A'-<{Singleton-singleton arcs remain the same, arc weight involving block entities are decided
          by an alpha density criteriontwork
          G=(V', E')
       End
   End
```

Social echelon – Partial Order Extraction

The second step is to convert the DAG to partial grades. That is to obtain a partition of the vertex in the DAG, $(V_1, V_2, \cdots, V_k)$, such that if $\overrightarrow{xy} \in \vec{A}$, then $x \in V_i$ and $y \in V_j$, where $i < j$ for some i, j. If there exists a link between a vertex in V_i and a vertex in V_j ($i < j$), the direction of the edge must be pointing from V_i to V_j. In the context of the Retweet network, users in V_i retweet users in V_j, but not the other way around. Although the partial ordered partition is not unique, the relative positions of the users still convey important information about the underlying structure of this network.

The partition is obtained by recursively removing all the “source” vertices – vertices with 0 in-degree, and assigning them to s. The algorithm generates a partition of 8 subsets. The result is shown in Appendix 1[7].

Social Influence Score

To test whether such echelon reflects or contains information about the social structure, or if it is just coincidental in this dataset, we construct an approximation of the user’s social influence to compare with the extracted echelon.

A number of researches have looked at the measure of influence in Twitter networks. (Kwak et al. 2010, Cha et al.2010, Bakshy et al. 2011) Popular choices of influence measures includes the number of followers, number of retweets, number of mentions, and PageRank score. Instead of using just one measure, we use the first dimension Principal Component Analysis score of the chosen measure. This one-dimension score is the best one-dimension representation of the measure data. Since we extract our echelon from the retweet network, we choose to use a set of measures that are independent of the retweet network, including Friendship relation authority score, total number of followers, Friendship relation in-degree, and Mention in-degree.

- Friendship relation authority score: Hub/Authority is a recursive ranking procedure that was originally designed to find highly endorsed websites, the “authorities”, and highly valuable lists that endorse other websites, the hubs, using link analysis algorithm. (Easley et al. 2010) (Kleinberg 1998) The Friendship relation between Twitter users resembles that between websites. A Twitter user automatically generates a list of “friends” that he/she chooses to follow. The list is visible to other users in the community, and serves as the endorsement of that user. It is likely that the users that follow a lot of important people have better ideas about where the prominent people are. The users who follow highly reputable users will get higher score as a hub. The users that are followed by high hub-score users will get higher score as authority. Only the authority score is included as a measure because it indicates the prominence of a user. Since Mention relation represents dialog rather than endorsement, its hub/authority score will not be included.
- Number of followers: The second factor that we take into account is the user’s total number of followers. A user’s number of followers is the most straightforward index of the user’s social influence in the Twitter community.
- In-degree of Friendship Relation: While the total number of followers score measures a user’s popularity in the entire Twitter network, the in-degree of the Friendship relation shows the user’s popularity within this sub-community.
- Number of mentions: Although we cannot infer the social status of a user just by looking at who replies to his/her message, the response rate of one’s tweets can reflect the importance of the user in that people usually pay more attention to messages sent from a prominent user. Table 4 below lists the network level in- and out- degree centralization of each relation. Notice the high in-degree centralization of the *Mention* network. It indicates the high heterogeneity of actor in degree centrality in the network. Comparing with the relatively low out degree centralization, it suggests that although people reply to other users’ messages quite evenly, only a few key users get more responses. Besides, in the technology enthusiasts’ community, oftentimes, it is the prestigious users who have access to new and groundbreaking tech news stories that fuels discussion.

Centralization	Friend (G1)	Mention (G2)	Retweet (G3)
Avg. Out degree	23.77806	11.94968	2.073548
Avg. Out degree of those with out-degree >=1	38.7958	25.2678	6.7014
In Degree Centralization	0.2112	0.4956	0.1681
Out Degree Centralization	0.237	0.1205	0.0879

[7] Full result can be found in the data packet.

Table 2: Network Overview

The table below shows the mean and standard deviation of social status scores in the eight echelons respectively. It is to be noted that since the partial order partition from DAG is not unique and our partial order extraction algorithm doesn't optimize for the best representation, which may cause the high standard deviation. Nevertheless, with the exception of V3, the echelons and the social scores match quite well.

	Mean	Standard Deviation
V8	55.37	65.63
V7	39.42	56.65
V6	20.95	81.90
V5	14.40	43.88
V4	1.18	29.38
V3	11.01	41.64
V2	2.39	40.28
V1	-13.41	17.17

Table 3: Social Scores of Each Echelons

CONCLUSION AND FUTURE WORK

In this research, our major finding is that in this cohesive Twitter sub-community the retweet behaviors among users are affected by the underlying echelons of social influence. The matching between the extracted echelons and the social influence shows that the presence of such echelons is not a transient phenomenon since it is mapped with the long-term measures of influence such as number of followers, authority score of friendship relation and so on. This property has not been reported in researches on generic Twitter Networks. However, we acknowledge that our dataset is relatively small and consists of only one community, which leaves room for further work. As the next step, we attempt to identify overlapping communities in Twitter network, and study if influence structure can be found in these communities. The result of this research also suggests that even in an informal, organic community such as the online technology enthusiasts community, semi-hierarchical social structure also presents itself, which indicates some similarity between online and offline community.

REFERENCES

1. Java, A., Song, X., Finin, T. and Tseng, B. (2007) Why we twitter: understanding microblogging usage and communities *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis,* 56-65.
2. Lancichinetti, Fortunato, S. and Kert´esz, J. (2009) Detecting the overlapping and hierarchical community structure in complex networks. *New Journal in Physics*, 11, 3
3. Bakshy, E., Hofman, J. M., Mason, W. A. and Watts, D. J. (2011) Everyone's an influencer: quantifying influence on twitter. *Proceedings of the fourth ACM international conference on Web search and data mining:* 65-74. Hong Kong, China: ACM.
4. Bollobas, B. (1998). *Modern Graph Theory*. Springer.
5. Watts, D and Dodds, P. (2007) Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*.
6. Easley, D. and Kleinberg, J. (2010) Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press.
7. Keller, E. and Berry, J. (2003.) The Influentials: One American in Ten Tells the Other Nine How to Vote, Where to Eat, and What to Buy. New york, NY: Free Press.
8. Wasserman, S. and Faust, K. (1994) Social Network Analysis: Methods and Applications. New York: Cambridge University Press.
9. Huberman, B., Romero, D. and Wu, F (2008) Social networks that matter: Twitter under the microscope. arXiv:0812.1045v1

10. Karp, R. M. (1972) Reducibility Among Combinatorial Problems. In R. E. Miller & J. W. Thatcher (Eds.) Complexity of Computer Computations: 85-103. Plenum Press.
11. Kleinberg, J. (1998) Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46:5, 604-32.
12. Kleinberg, J. (2008) The convergence of social and technological networks. *Communications of the ACM*, 51:11, 66–72.
13. Krackhardt, D. (1994) Graph Theoretical Dimensions of Informal Organizations. In *Computational Organization Theory* (Ed.).
14. Krishnamurthy, B., Gill, P. and Arlitt, M. (2008) A few chirps about twitter. *Proceedings of the first workshop on Online social networks*, 19-24
15. Kwak, H., Lee, C., Park, H. and Moon, S. (2010) What is Twitter, a social network or a news media? *Proceedings of the 19th international conference on World wide web:* 591-600. Raleigh, North Carolina, USA: ACM.
16. Cha, M., Haddadi, H., Benevenuto, B. and Gummadi, P. (2010) Measuring user influence on twitter: The million follower fallacy. In *4th Int'l AAAI Conference on Weblogs and Social Media*,.
17. Domingos, P. and Richarson, M. (2001) Mining the Network Value of Customers. *Proceeding of the Seventh International Conference on Knowledge Discovery and Data Mining.*
18. Richarson, M. and Domingos, P. (2002) Mining Knowledge-Sharing Sites for Viral Marketing. *Proceeding of the Eighth International Conference on Knowledge Discovery and Data Mining.*
19. Rogers, E. 1962. Diffusion of Innovations. Free Press.
20. Weng, J., Lim, E., Jiang, J. and He, Q. (2010) TwitterRank: finding topic-sensitive influential twitterers. *Proceedings of the third ACM international conference on Web search and data mining:* 261-70. New York, New York, USA: ACM.

APPENDIX 1

Echelon(# of users)	List of User IDs
V_8 (6)	'816653' '652193' '972651' '6273552' '11348282' '13'
V_7 (11)	'57203' '783214' '2172' '8453452' '809760' '30863' '94143715' '586' '20' '14321959' '44570946'
V_6 (28)	'36823' '1051171' '2729061' '1835951' '12528' '414' '15738725' '14348594' '15661871' '33423' '30313925' '817386' '37570179'….
V_5 (82+2)	'61133+5905672' '819606+815973' '820585' '16953157' '6735' '418' '5637652' '30331417' '732073' '18327902' '11113' '1422311' '817268'….
V_4 (10+173)	'12+19002481+6141832+16895951+1344951+2713951+3471+12089102+648+….' '6503412' '21879024' '5676102' '82788404' '14331688' …
V_3 (36)	'3829151' '618593' '12019742' '14712874' '12514' '41783' '663463' '755859' '644603' '33923' '7846' '6897142' '14334532'….
V_2 (90+1)	'12101862+14980437' '10938882' '6253282' '11661' '678953' '9184282' '13461' '823083' '10202' '13479' '10990'….
V_1 (153+2)	'15827269+9641832' '11231232+12741' '19225408' '6297382' '17218144' '98035778' '799722' '7090182' '12565032' '14492722' '755431' '765694' …